Developmental Biology 383 (2013) 158–173

Contents lists available at ScienceDirect

Developmental Biology

journal homepage: www.elsevier.com/locate/developmentalbiology

Shadow enhancers flanking the *HoxB* cluster direct dynamic *Hox* expression in early heart and endoderm development

Christof Nolte [a], Tim Jinks [b,1], Xinghao Wang [a,c,2], María Teresa Martinez Pastor [b,3], Robb Krumlauf [a,b,c,*]

[a] Stowers Institute for Medical Research, Kansas City, MO 64110, USA
[b] National Institute for Medical Research, London NW7 1AA, United Kingdom
[c] Department of Anatomy and Cell Biology, Kansas University Medical Center, Kansas City, KS 66160, USA

ARTICLE INFO

Article history:
Received 14 June 2013
Received in revised form
3 September 2013
Accepted 11 September 2013
Available online 18 September 2013

Key words:
Hox genes
Heart
Retinoids
Enhancers
Gene regulation
Endoderm

ABSTRACT

The products of *Hox* genes function in assigning positional identity along the anterior–posterior body axis during animal development. In mouse embryos, *Hox* genes located at the 3′ end of *HoxA* and *HoxB* complexes are expressed in nested patterns in the progenitors of the secondary heart field during early cardiogenesis and the combined activities of both of these clusters are required for proper looping of the heart. Using *Hox* bacterial artificial chromosomes (BACs), transposon reporters, and transgenic analyses in mice, we present the identification of several novel enhancers flanking the *HoxB* complex which can work over a long range to mediate dynamic reporter expression in the endoderm and embryonic heart during development. These enhancers respond to exogenously added retinoic acid and we have identified two retinoic acid response elements (RAREs) within these control modules that play a role in potentiating their regulatory activity. Deletion analysis in *HoxB* BAC reporters reveals that these control modules, spread throughout the flanking intergenic region, have regulatory activities that overlap with other local enhancers. This suggests that they function as shadow enhancers to modulate the expression of genes from the *HoxB* complex during cardiac development. Regulatory analysis of the *HoxA* complex reveals that it also has enhancers in the 3′ flanking region which contain RAREs and have the potential to modulate expression in endoderm and heart tissues. Together, the similarities in their location, enhancer output, and dependence on retinoid signaling suggest that a conserved *cis*-regulatory cassette located in the 3′ proximal regions adjacent to the *HoxA* and *HoxB* complexes evolved to modulate *Hox* gene expression during mammalian cardiac and endoderm development. This suggests a common regulatory mechanism, whereby the conserved control modules act over a long range on multiple *Hox* genes to generate nested patterns of *HoxA* and *HoxB* expression during cardiogenesis.

Introduction

Hox genes code for transcription factors whose differential expression during early embryogenesis assigns positional identity along the body axis. In mammals, there are 39 genes organized into four complexes (A, B, C, and D) with each complex containing from 9 to 11 genes (Krumlauf, 1994). There is an association between their chromosomal organization and sequential activation known as colinearity (Duboule and Dollé, 1989; Graham et al., 1989; Kmita and Duboule, 2003). This regulatory feature is used in many tissues and organs to establish a nested complement of Hox proteins along embryonic axes that forms part of a molecular code for specification of positional information (Alexander et al., 2009; Zakany and Duboule, 2007; Mallo et al., 2010; Wellik, 2009). Retinoic acid (RA) signaling plays an important role in establishing the initial patterns of expression of *Hox* genes in the neural ectoderm, mesoderm and endoderm (Alexander et al., 2009; Deschamps and van Nes, 2005; Huang et al., 1998; Tümpel et al., 2009). Experimental manipulation of RA levels or inactivation of components involved in RA signaling and metabolism result in changes in patterns of *Hox* expression during early development (Rhinn and Dolle, 2012; Sirbu et al., 2005). For some of these *Hox* genes, input from RA signaling is directly mediated by

* Corresponding author at: Stowers Institute for Medical Research, 1000 East 50th Street, Kansas City, MO 64110, USA.
E-mail addresses: t.jinks@wellcome.ac.uk (T. Jinks), xwang@hsml.com (X. Wang), Maria.Teresa.Martinez@uv.es (M.T. Martinez Pastor), rek@stowers.org (R. Krumlauf).
[1] Current address: The Wellcome Trust Limited, Gibbs Building 215 Euston Road, London NW1 2BE, United Kingdom.
[2] Current address: HSML, 45 South Seventh Street, Suite 2700, Minneapolis, MN 55402, USA.
[3] Current address: Department de Bioquímica i Biologia Molecular, Universitat de València, Dr. Moliner 50, 46100 Burjassot, Spain.

0012-1606/$ - see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.ydbio.2013.09.016

cis-regulatory sequences known as retinoic acid response elements (RAREs) (reviewed in Tümpel et al. (2009)).

The heart is one of the first organs to develop in the vertebrate embryo (Buckingham et al., 2005; Vincent and Buckingham, 2010). Its construction originates from four different regions of the developing embryo. Two of these regions, the first and second heart field are derived from the splanchnic mesoderm. The first heart field (FHF) gives rise to the heart tube and subsequently the left ventricular myocardium. Cells from the second heart field (SHF) use the heart tube as a scaffold in which to migrate and contribute to the myocardium of the atria, right ventricle and outflow tract (OFT). In addition to these two mesodermally-derived fields, cells from the cardiac neural crest and the proepicardial organ (PEO) make contributions. The cardiac neural crest cells contribute to the musculature of the arteries connected to the heart and are required for the correct morphology and septation of the cardiac outflow tract in birds and mammals (reviewed in Creazzo et al. (1998) and Stoller and Epstein (2005)). The PEO is a temporary mesenchymal population of cells derived from the septum transversum, located posterior to the developing heart. Cells from the PEO migrate and form the outer lining of the heart, the epicardium.

Endoderm plays a crucial role during cardiogenesis, both as a physical substrate over which cells of the cardiogenic mesoderm migrate, as well as a source of instructive factors that regulate the cardiac specification and differentiation program (Dunwoodie, 2007). Both early endoderm progenitors and pre-cardiac mesoderm are specified in close proximity to one another within the anterior region of the primitive streak (Varner and Taber, 2012). Concurrently early *Hox* gene expression is initiated in the cells ingressing through the node (Deschamps and van Nes, 2005). As development proceeds, collinear *Hox* gene expression becomes regionalized during endoderm development and correlates with local features of the developing gut and genitalia (Dollé et al., 1991; Grapin-Botton, 2005). Loss- and gain-of-function studies of *Hox* genes have demonstrated roles in patterning a variety of endodermal tissues, such as the thymus, esophagus, stomach, intestine and genitalia and both the ileo-caecal and anal sphincters (Aubin et al., 2002; Boulet and Capecchi, 1996; Kondo et al., 1996; Manley and Capecchi, 1998; Wolgemuth et al., 1989; Zacchetti et al., 2007; Zakany and Duboule, 1999; Zákány et al., 2001).

Hox genes are expressed throughout the splanchnic mesoderm, rhombomeres (origin of cardiac neural crest cells), and the septum transversum – tissues from which progenitors of the heart originate. *Hox* gene expression in the heart was first characterized in the chick embryo, where *Hoxd3*, *Hoxa4* and *Hoxd4* were detected during early cardiogenesis (Searcy and Yutzey, 1998). Subsequently, *Hoxa1*, *Hoxb1* and *Hoxa3* have been shown to be expressed in the cardiac progenitors of the mouse heart (Bertrand et al., 2011) where their descendants populate the atria and OFT. These nested expression patterns suggest that multiple Hox proteins may modulate aspects of the cardiac developmental program. However, few cardiac defects have been reported as a result of *Hox* mutations. In the mouse, targeted disruption of the *Hoxa3* gene results in heart and artery defects, primarily due to changes in the fate of the cranial neural crest cells that normally express *Hoxa3* (Chisaka and Capecchi, 1991). Mutations in the *HOXA1* gene result in human cardiovascular defects that include malformations of the OFT and the arteries connected to the heart (Tischfield et al., 2005). Recently, a re-evaluation of the *Hoxa1* mouse mutant revealed phenotypes that closely model those observed in the human *HOXA1* syndrome patients (Makki and Capecchi, 2012). These studies highlight roles for the *Hoxa1* and *Hoxa3* genes in patterning progenitors arising from the contribution of the cardiac neural crest cells to the heart. The absence of significant heart phenotypes in single targeted mutants appears to be due to functional compensation by other genes or clusters. Deletion of either the *HoxA* or *HoxB* complexes does not result in enhanced heart phenotypes compared to the individual gene mutations. In contrast, the specific compound deletion of both the *HoxA* and *HoxB* complexes results in mouse embryos that fail to undergo proper heart looping (Soshnikova et al., 2013). Other variants of double cluster deletions do not present these heart defects, revealing a specific combined role for multiple members from the *HoxA* and *HoxB* complexes in heart morphogenesis.

Heart looping defects are seen in mouse embryos deficient in RA signaling. In the absence of functional retinaldehyde dehydrogenase 2 (Raldh2), the enzyme that converts retinaldehyde into retinoic acid, the heart tube forms but fails to undergo looping (Niederreither et al., 1999, 2001). In *Raldh2* null mutant embryos, the gene expression profile of the SHF is altered suggesting a posterior expansion of the field (Ryckebusch et al., 2008). The role of RA signaling in determining the size of the heart field and regulating the position of the posterior domain of the cardiac field is conserved in other vertebrate models (Yutzey et al., 1994). The influence of RA on heart development may be partially mediated through its effects on *Hox* genes. RA signaling has an indirect role in determining the size of the cardiac field mediated by its regulation of *Hoxb5b* expression in the adjacent forelimb field (Keegan et al., 2005; Waxman et al., 2008). In the absence of RA, there is reduction in the forelimb field and a posterior expansion of the cardiac field, revealing retinoid dependent communications between these adjacent fields. In the mouse, *Raldh2* null mutant embryos *Hoxa1* is no longer expressed in the SHF (Ryckebusch et al., 2008). Recently, an RA responsive enhancer for the *Hoxa3* gene has been characterized and shown to direct *lacZ* expression in cardiac neural crest cells and the SHF (Diman et al., 2011). RA signaling is also important for regionalization of the endoderm and formation of the pancreas (Kinkel et al., 2009; Stafford et al., 2006).

Together these studies suggest that RA-directed expression of *Hox* genes is important for patterning and morphogenesis of the endoderm and heart. However, with the exception of the murine *Hoxa3* gene (Diman et al., 2011), no enhancers capable of mediating *Hox* expression in the progenitors of the developing heart have been identified. Similarly, few regulatory regions for the directed expression of *Hox* genes in endoderm derivatives have been characterized (Huang et al., 1998). Here we report the identification of several novel enhancers from the *HoxA* and *HoxB* complex that mediate reporter expression in endoderm and regions that are sources of cardiac progenitors, including the proepicardial organ (PEO) and splanchnic mesoderm, as well as in the OFT and epicardium of the embryonic heart. These enhancers contain RAREs, suggesting that they integrate inputs from retinoid signaling while modulating *HoxA* and *HoxB* expression during heart development and endoderm patterning.

Materials and methods

Comparative genomics analysis

Comparative genomic analysis was performed at the mVISTA website using the LAGAN alignment tool (http://genome.lbl.gov/). The genomic sequences used in the alignments were downloaded from the Ensembl website (http://www.ensembl.org/) and imported into VectorNTi.

Transposon insertions in HoxB BACs

The technique used to generate the *lacZ* reporter transposon studded *HoxB* BACs has been previously described (Morgan et al.,

1996). The original host strain, HB101 (endA$^+$), was substituted with HB10B as it resulted in BAC DNA instability. The BAC used in this study (MMP-5) was isolated from a 129SV mouse BAC library (Research Genetics, USA; for further details please see Parrish et al. (2011)). Reporter integrations were initially screened by alkaline lysis mini preparatory analysis, taking advantage of an internal *Sma*I site within the *lacZ* transposon to identify BAC clones in which the transposon had integrated downstream of the *HoxB* complex. The actual site of transposon integration was identified by sequencing outwards from the 5′ and 3′ ends of the integrated transposon using the γ and Δ oligonucleotides described in Morgan et al. (1996).

BAC deletion constructs

Two different types of deletions were generated using the recombinogenic bacterial strains, EL250 and EL350, from the Neal Copeland laboratory (Lee et al., 2001). Deletion of the 3′ *Hoxb1* enhancers was accomplished by using a replacement strategy such that these elements were exchanged for an Frt flanked kanamycin selection cassette (derived from pIGCN21). Once positive clones were identified, the kanamycin selection cassette was removed by growing in 3 ml LB cultures supplemented with 0.1 M sucrose for 1 h at 32 °C. In the second type of deletion, the novel regions downstream of the *HoxB* complex that displayed enhancer activity were deleted using a loxP flanked kanamycin–neomycin cassette.

Capture constructs and generation of successive 3′ truncations

To generate the 10 kb reporter constructs and truncated variations, oligonucleotide cassettes were designed to 'capture' the region of interest from a BAC spanning the region between *Hoxb1* and *Skap1* into pBGZ40 (Yee and Rigby, 1993) using recombinogenic bacterial strain DY380 (Lee et al., 2001). These oligonucleotide cassettes contain a 5′ and 3′ 47 nucleotide long homology arm corresponding to the flanking ends of the region to be 'captured' separated by an unique *Sal*I site. This 'capture' technique did not work to clone Region 6 as we could not propagate the clone in bacteria. Additional approaches, such as PCR amplification and direct subcloning from the BAC, also failed to generate a propagable plasmid containing Region 6.

Preparation of DNA for microinjections, transgenesis and β-galactosidase staining assays

The BAC DNA from positive clones was prepared using the QIAGEN Maxi prep kit with the following modifications. Bacteria from a 300 ml overnight culture were pelleted and resuspended in 30 ml of P1 buffer. The volumes of the subsequent buffers (P2 and P3) were each increased to 30 ml, while the incubation time following the addition of P3 was reduced to 10 min. BAC DNA was eluted from the column with 65 °C warmed QF solution.

BAC DNA was prepared for microinjection by loading onto a QIAquick PCR purification kit and then eluting it in 50 °C prewarmed microinjection buffer. Following quantitation, it was subsequently diluted in BAC microinjection buffer to a concentration of 1–2 ng/μl for pronuclear microinjections. Plasmid based reporter fragments were separated from vector sequence by gel electrophoresis and the inserts extracted from agarose using MinElute (Qiagen). For microinjections they were prepared using the same method used for BAC DNA except they were resuspended in microinjection buffer without the addition of spermidines. Transgenic embryos and founders were generated by pro-nuclear injection of linearized constructs into fertilized eggs of C57Bl/10J xCBA-F1 crosses and reimplanted into pseudopregnant foster mothers.

To detect β-galactosidase activity, embryos were fixed in either 0.1% paraformaldehyde/0.2% glutaraldehyde (E11.5–E13.5) or 4% paraformaldehyde (PFA) (E14.0 or older) for 30–60 min on ice. After several washes in phosphate buffered saline, samples were stained in X-Gal for 4–20 h at 4 °C or at room temperature (Ahn et al., 2013; Whiting et al., 1991). The frequency of transgenic embryos was determined by PCR reactions using primers against the *lacZ* sequence of the β-galactosidase reporter.

Retinoic acid gavage

Pregnant transgenic dams were feed all-trans retinoic acid (Sigma Aldrich, R2625) first dissolved in DMSO and then diluted in sesame seed oil with a final concentration of 20 mg/kg of body weight. Treatment times were performed 6 h prior to collecting E7.5 and E9.5 embryos for *lacZ* expression assays.

Mutation of RARE sequences

The novel RARE sequences were mutated by PCR overlap extension (Ho et al., 1989). Oligonucleotides were designed such that the direct repeat nucleotides were converted to adenines. Clones were sequenced for confirmation of mutation of the RARE sequence before being ligated back into the *lacZ* reporter plasmid and prepared for pronuclear microinjections.

Results

The goal in this study was to identify regulatory regions which could participate in the regulation of *HoxB* and *HoxA* genes in heart and endoderm development. Our previous regulatory analyses of the *Hox* complexes uncovered a variety of local embedded enhancers that mediated restricted expression in a variety of tissues (Tümpel et al., 2009) but none within the developing embryonic heart. It is possible that global control regions or enhancers working at long range, analogous to those for *HoxD* expression in the limb (Zakany and Duboule, 2007), may be involved in mediating heart expression; therefore, we explored the *cis*-regulatory potential of the 3′ flanking intergenic regions adjacent to the most anteriorly expressed genes.

Comparative analysis of the region flanking the HoxB complex between placental mammals

Phylogenetic footprinting is a useful tool to search for potential enhancer regions based on sequence conservation. Analysis with the VISTA comparative genomics program, using mouse as a reference sequence, revealed numerous blocks of conservation spanning the entire intergenic region between *Hoxb1* and *Skap1* (Fig. 1), its next non-*Hox* neighbor gene whose expression is restricted to T-lymphocytes, mast cells and macrophages (Wang and Rudd, 2008). These blocks of sequence varied in size and location but were unique to placental mammals as few of them were conserved when the mouse was used a reference sequence against representative non-placental vertebrates, such as turtle and fish (Fig. 1C).

The first 5 kb of sequence at the 5′ end of this intergenic interval displays multiple peaks of conservation which correspond to three previously identified local enhancers of the *Hoxb1* gene. These are located 3′ of the gene and are responsible for its expression in the neural tube, mesoderm and foregut endoderm (Gavalas et al., 1998; Huang et al., 1998; Marshall et al., 1994; Studer et al., 1998). We observed a similar level of broad conservation spread throughout the entire intergenic region between *Hoxb1* and *Skap1* (Fig. 1B). These blocks of conserved sequence

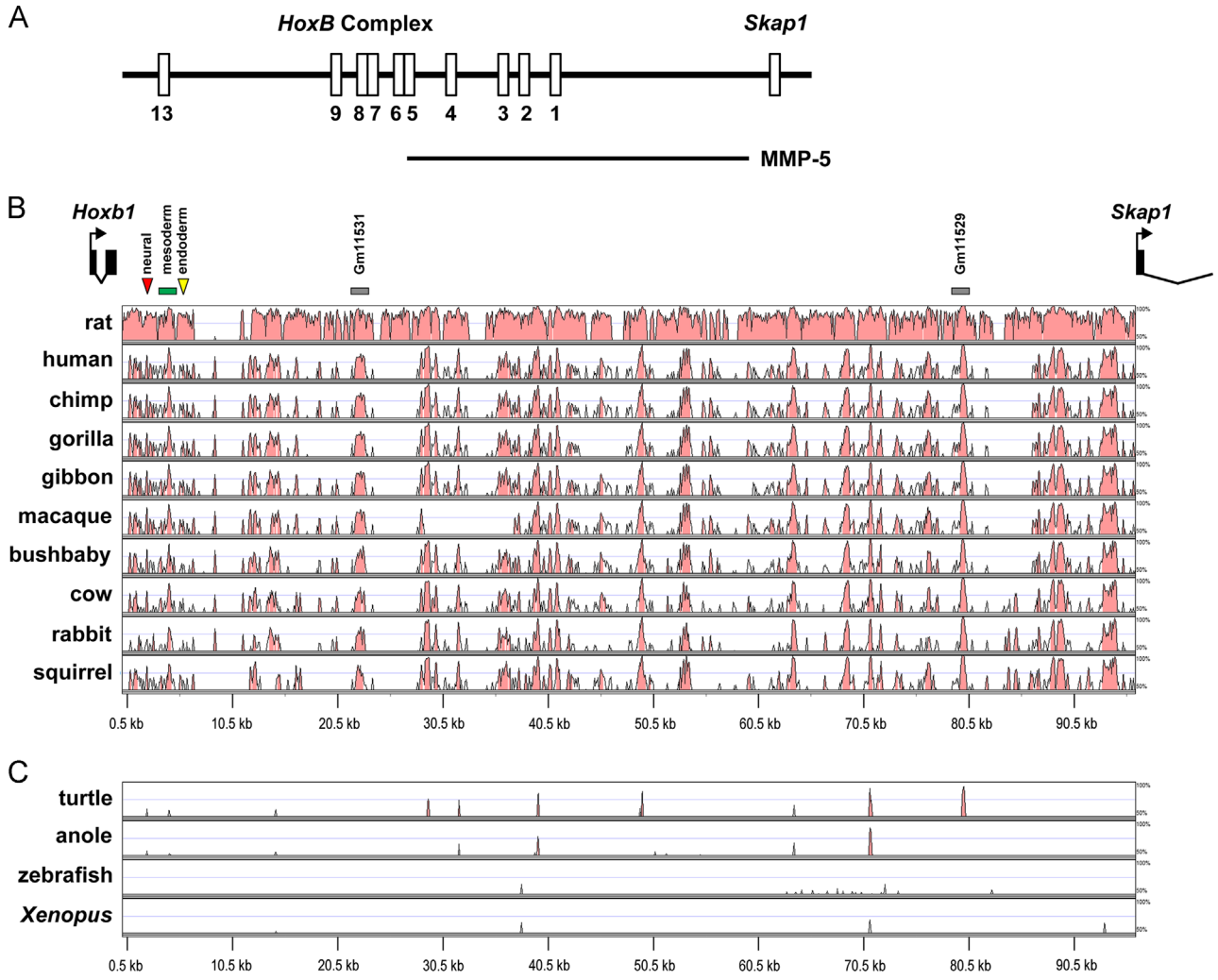


Fig. 1. VISTA analysis of the intergenic region between *Hoxb1* and *Skap1*. (A) A genomic map illustrating the relationship between the *HoxB* complex and the first exon of the *Skap1* gene. Below it, the approximate range of the BAC (MMP-5) modified and used in this study to examine the regulatory potential of the region between *Hoxb1* and *Skap1*. (B) Comparison of intergenic sequences from available placental mammals, using the mouse as the reference sequence, show high level of sequence conservation within the region between *Hoxb1* and *Skap1*. (C) This level of sequence conservation is not observed when a similar VISTA comparison of non-mammal vertebrates is performed against the mouse. In (B) the known local enhancers of the *Hoxb1* gene that drive its expression during development are indicated as a redhead arrow (neural tube enhancer), green box (mesoderm/endoderm enhancer) and yellow arrowhead (foregut endoderm enhancer). Also illustrated, as gray boxes above the placental vertebrate comparison, are a pseudogene (Gm11531) and the 3′ end of a noncoding RNA (Gm11529).

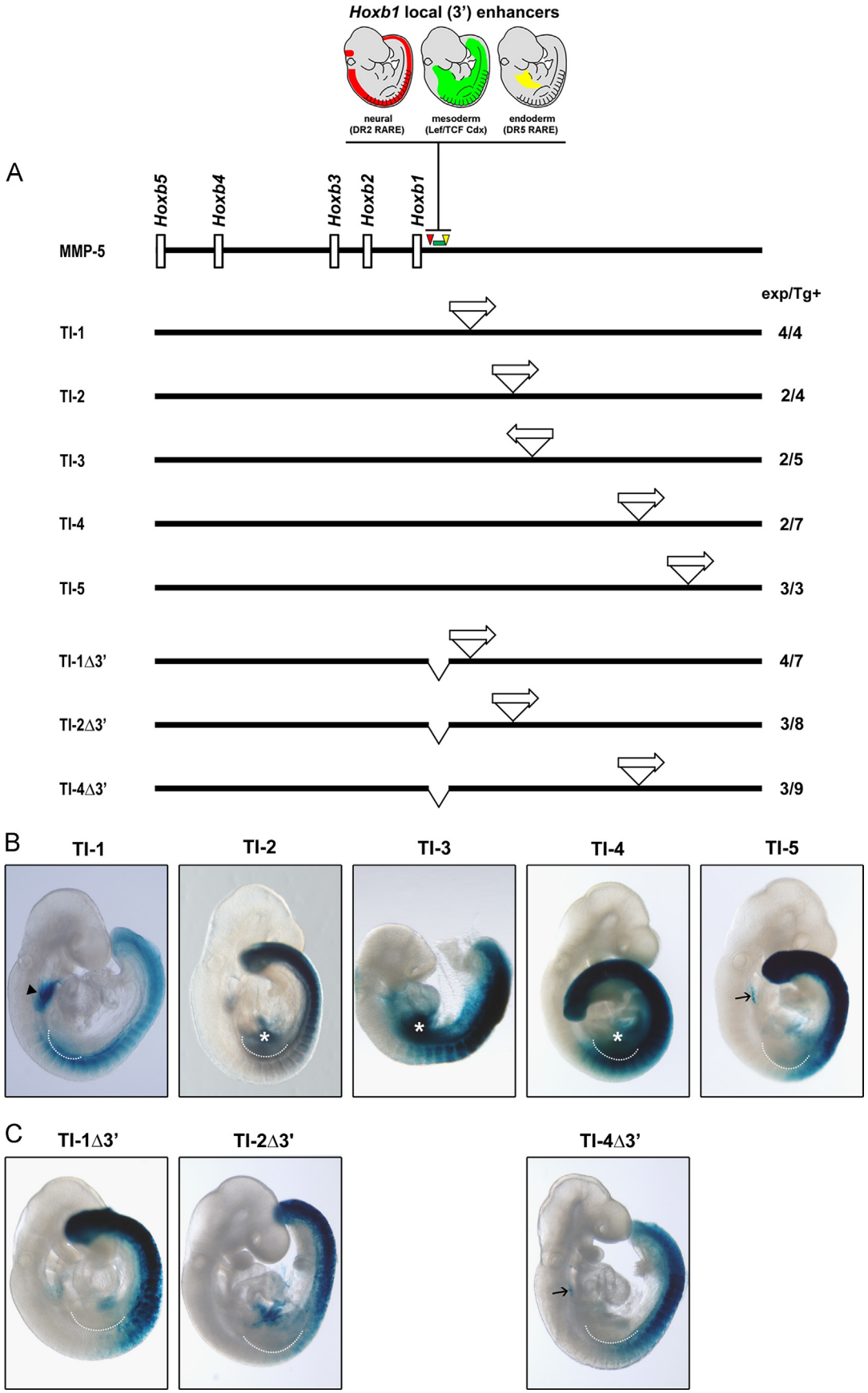
could correspond to several types of elements such as enhancers, pseudogenes, long noncoding RNAs, scaffold/matrix attachment regions, insulators and boundary elements maintained during the mammalian lineage. In fact, the conserved blocks at 22 and 80 kb correspond to a known pseudogene (Gm11531) and the 3′ end of long noncoding RNA (Gm11529), respectively. In light of this extensive conservation throughout the intergenic region it was not possible to focus on any particular area as a candidate for regulatory activity in the heart and endoderm. Therefore, it was necessary to utilize alternative approaches to functionally test the regulatory potential within this intergenic region.

Transposon-reporter assay of the region flanking the HoxB complex

To assay the regulatory potential of this region at different intervals we generated a library of transposon insertions (TIs) into a BAC clone (MMP-5) that contained the 3′ half of the *HoxB* complex (i.e., *Hoxb5* to *Hoxb1*) as well as 80 kb of the flanking intergenic region downstream of the *Hoxb1* gene (Fig. 1A). The transposon contained a *lacZ* reporter under the control of a β-globin minimal promoter (serving as an enhancer trap) and randomly integrated as a single copy once per BAC molecule (Morgan et al., 1996). Precise integration sites were determined and several clones that contained the *lacZ*-reporter transposed at 10 kb (TI-1), 21 kb (TI-2), 26 kb (TI-3), 52 kb (TI-4) and 64 kb (TI-5) downstream of the *Hoxb1* gene were chosen for transgenic analysis (Fig. 2A).

The transposon reporter integrated closest to the *HoxB* complex, TI-1, displayed strong *lacZ* expression in the presumptive foregut region, neural epithelium and somitic mesoderm (Fig. 2B) in a pattern predicated from transgenic analyses of the previously characterized individual 3′ enhancers of *Hoxb1* (Gavalas et al., 1998; Huang et al., 1998; Marshall et al., 1994; Studer et al., 1998). This indicated that the TI-1 transposon was successfully integrating regulatory influences from these known local *Hoxb1* enhancers.

Despite differences in the integration site or orientation of the reporter the overall expression pattern in E9.5 transgenic mouse embryos of the other four transposon insertions (TI-2 to TI-5) was similar, with *lacZ* expression throughout the posterior trunk (Fig. 2B). Reporter expression was strong throughout the presomitic mesoderm of the tail bud and became progressively restricted to the ventral aspect of the somites, extending in TI-1 to TI-4 up to, and


Hoxb1 local (3') enhancers
neural (DR2 RARE)
mesoderm (Lef/TCF Cdx)
endoderm (DR5 RARE)
A
Hoxb5
Hoxb4
Hoxb3
Hoxb2
Hoxb1
MMP-5
exp/Tg+
TI-1
4/4
TI-2
2/4
TI-3
2/5
TI-4
2/7
TI-5
3/3
TI-1Δ3'
4/7
TI-2Δ3'
3/8
TI-4Δ3'
3/9
B
TI-1
TI-2
TI-3
TI-4
TI-5
C
TI-1Δ3'
TI-2Δ3'
TI-4Δ3'

including, the territory adjacent to the forelimb bud. In TI-5, this anterior domain was slightly posteriorized. Although the overall pattern was similar between the TI reporters, there were important reproducible differences between the sites of integration. BAC reporters TI-2, TI-3 and TI-4 produced strong expression in the septum transversum (Fig. 2B, white asterisk) and visceral endoderm which was not observed with TI-1. In addition, TI-3 and TI-4 displayed β-galactosidase activity throughout the visceral mesoderm whereas in TI-2 expression in the visceral mesoderm was restricted to the domain in the septum and absent from the posterior regions. In TI-4 and TI-5, a small group of *lacZ* expressing cells lining the dorsal aorta was observed (Fig. 2B, black arrows).

The similarities and differences in the patterns of reporter expression of these transposons suggest that they maybe integrating regulatory influences from known enhancers and other elements spread over a broad region flanking the *HoxB* cluster. The common overall pattern of expression shared by these transposons could be the result of the known local 3′ *Hoxb1* enhancers acting over longer distances to mediate *lacZ* expression. To directly evaluate this possibility, we deleted a 4.4 kb region encompassing the local *Hoxb1* 3′ enhancers from three of the TI BAC constructs to assay its input on expression (Fig. 2A). In all three lines with this deletion, the major domains of reporter expression remained remarkably similar (Fig. 2C). While expression in the tail bud presomitic region remained strong, the somitic domain of expression was weakened and posteriorized such that its anterior border aligned to the caudal side of the developing forelimb bud. Expression in the visceral mesoderm, endoderm and the septum transversum was greatly reduced or abolished. The foregut domain of *lacZ* expression present in TI-1 (Fig. 2B, black arrowhead) was significantly reduced in TI-1Δ3′ indicating that this domain of expression was primarily generated by the *Hoxb1* 3′ endodermal enhancer (Huang et al., 1998) deleted in this construct. The changes in reporter expression arising from deletion of the 3′ proximal *Hoxb1* enhancers revealed that they are capable of exerting regulatory influences over a long range.

Identification of new endoderm and heart enhancers flanking HoxB

The observation that transposons bearing the deletion of the known elements displayed very similar expression patterns, suggested that additional regulatory modules with overlapping activities may be present in this *HoxB* flanking region. To screen for this potential, we divided the 80 kb of the intergenic region into a series of overlapping 10 kb constructs and tested these in transgenic *lacZ* reporter assays (Fig. 3A). Of the seven 10 kb reporters that we tested (Region 6 was unclonable), Regions 1–4, corresponding to the first 40 kb flanking the *Hoxb1* gene, displayed reproducible and independent patterns of expression (Fig. 3B–E). Region 1, which contains the previously characterized *Hoxb1* 3′ enhancers, produced the expected *lacZ* expression throughout the trunk mesoderm and foregut endoderm (Fig. 3B). Region 2 produced a small patch of *lacZ* stained cells in the visceral endoderm and proepicardial organ (PEO) (Fig. 3C, white arrowhead). The *lacZ* expression pattern of Region 3 included domains in the ectoderm of the branchial arches, the viscercal endoderm, the PEO, and the epicardium (Fig. 3D). Region 4 produced *lacZ* expression in the wall of the dorsal aorta (Fig. 3E, black arrow) and in a small patch of cells located in the ventral–medial aspect of the developing heart (Fig. 3E, open arrow).

This transgenic analysis revealed multiple regions with regulatory potential. Some of these corresponded to novel domains of expression found in the transposon scanning assay. For example, Regions 2 and 3 (Fig. 3C–D) displayed the visceral endoderm/septum transversum *lacZ* expression patterns seen in TI-2, TI-3 and TI-4 (Fig. 2B). However, by examining these as independent fragments we also observed patterns of reporter activity that were not detected in the context of the transposon assay. This suggests they may need to work in context with surrounding elements to mediate appropriate patterns of expression.

In an attempt to delineate where within Regions 2–4, enhancer activity was concentrated, a series of subclones were generated involving successive 5′ or 3′ truncations (Fig. 4). For Regions 2 and 3, deletions of any sequence abolished reporter activity suggesting that multiple *cis*-elements are required for their regulatory potential. In contrast, 3′ deletions of Region 4 revealed a small 5′ fragment (4A) with expanded regulatory potential compared to the entire 10 kb Region 4 (Figs. 4 and 5). We noted from the VISTA sequence alignment that there was significant conservation within 4A (Fig. 4A). Furthermore, this conservation extended into the 3′ sequences of Region 3 which when deleted (3B) abolished activity (Fig. 4B). Therefore, we generated another reporter construct containing the 3′ end of Region 3 and the 4A subclone, which we called HGE (Fig. 4). This combination of regions resulted in reporter staining that displayed different dynamics and patterns compared to Region 3 or 4A alone (Fig. 5).

Temporal dynamics of enhancer activity in endoderm and components of the embryonic heart

To monitor the patterns of expression mediated by these regulatory regions during development we generated transgenic reporter lines and characterized their expression in whole mount embryos (Fig. 5). Region 3 displayed robust *lacZ* expression in the visceral ectoderm as well as in the lateral mesoderm, but was not expressed in the somites, neural tissues or tail bud at E8.5 (Fig. 5). At this stage *lacZ* expression is apparent in the splanchnic mesoderm that contributes to the SHF (Fig. 5, black square brackets). By E9.5 Region 3 mediated staining in the ectoderm overlaying the branchial arches, the epicardium, the PEO, and visceral endoderm. This overall pattern was maintained through E10.5 but new staining was detected in some of the developing somites. By E11.5, there was an upregulation in the somites and new domains of *lacZ* expression appeared near the eye and in the presumptive future cranial suture.

We detected no expression in embryos at E8.5 in the Region 4 reporter line. However, Region 4 mediated expression in a small patch of cells located medial lateral to the developing heart and robust expression in the dorsal aorta from E9.5 to E10.5. At E11.5 reporter staining is significantly reduced and is absent at later stages. A smaller 5′ fragment of Region 4 (4A) did not direct *lacZ* expression in the dorsal aorta, but in contrast mediated expression in the epicardium of the heart from E9.5 to E11.5 (Fig. 5) with earlier expression at E8.5 appearing in the anterior intestinal

Fig. 2. Transposon-*lacZ* BAC reporter constructs and their expression patterns in transgenic mouse embryos at E9.5. (A) Several clones of the MMP-5 BAC with the integrated transposon-*lacZ* reporter at different positions downstream of the *Hoxb1* gene were used in transient BAC transgenic assays to assay the regulatory potential of this region. The position and orientation of the transposon-*lacZ* reporters are indicated as white arrows. The frequency of expressors over transgenic embryos is displayed to the right side of each BAC construct (exp/Tg+). The local 3′ enhancers of *Hoxb1* and their approximate size are indicated as red (neural), green (mesoderm/endoderm) and yellow (foregut endoderm) boxes. Depicted above them are representations of E9.5 mouse embryos with their expression pattern, and below them, their core regulatory sequence/transcription factor listed in brackets. (B) The overall pattern of *lacZ* expression in E9.5 embryos was similar regardless of the site of reporter integration. As a morphological landmark, the forelimb bud of each embryo (except for TI-3) is highlighted with a dotted, white line. In embryos bearing the TI-1 BAC, the black arrowhead points at the domain of foregut expression driven by the local *Hoxb1* DR5 RARE 3′ enhancer which is lost in embryos bearing the deletion of the area (TI-1Δ3′). In TI-5 and TI-4Δ3′ transgenic embryos, the black arrow points to *lacZ* expressing cells flanking the dorsal aorta. The white asterisk indicates the septum transversum. (C) Deletion of the known *Hoxb1* enhancers results in a posterior shift in the expression of the BAC transposon-*lacZ* reporters, but does not eliminate its activity.

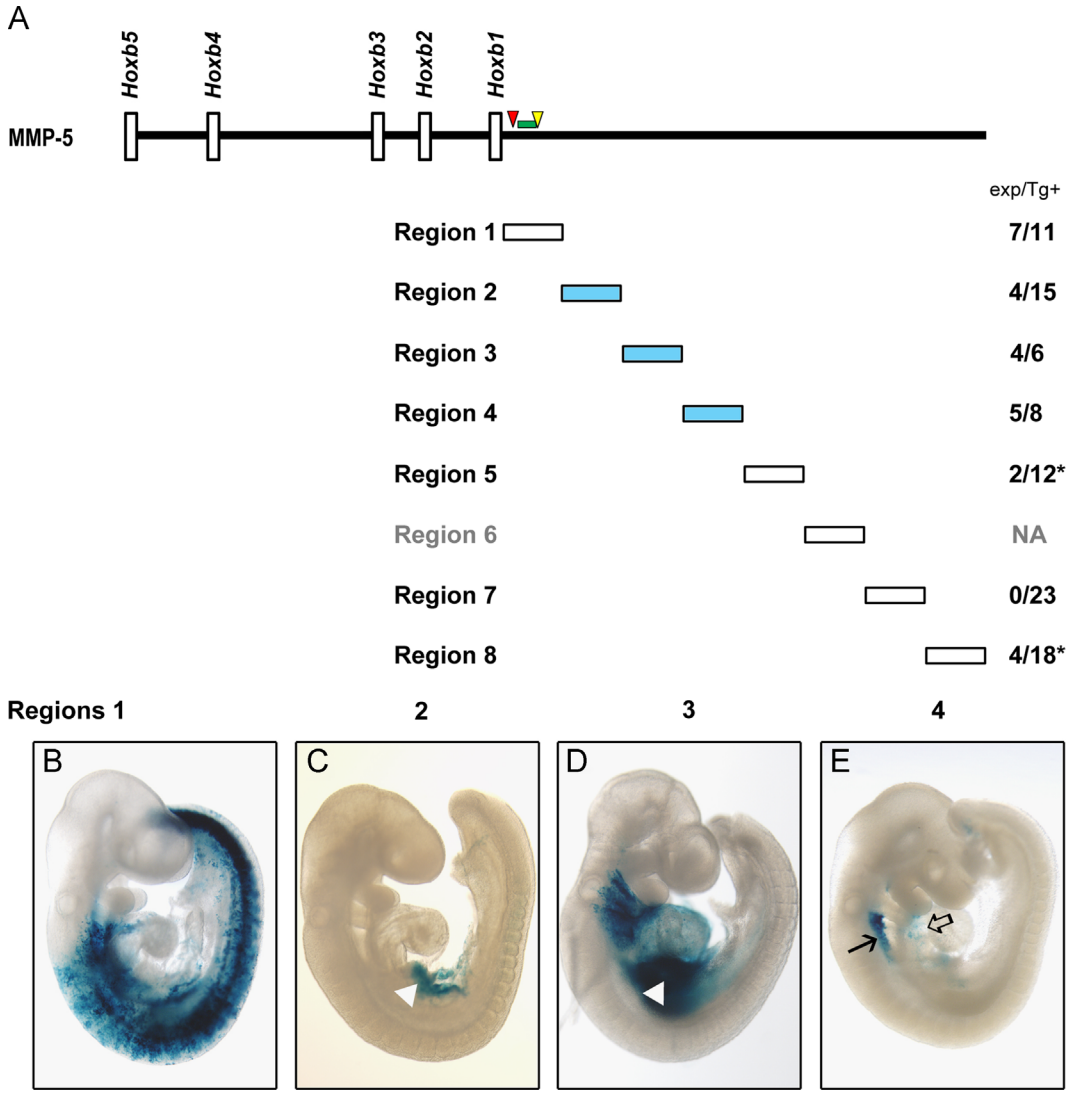


Fig. 3. Transgenic analysis of 10 kb subclones derived from the *Hoxb1-Skap1* intergenic region and their expression pattern in E9.5 mouse embryos. The 80 kb region immediately downstream of the coding sequence of the *Hoxb1* gene was divided into eight 10 kb Regions that overlapped each other by 200 nucleotides. Of these seven regions, only Regions 1–4 had reproducible patterns of *lacZ* expression (B–E). Embryos with Regions 5 and 8 had unreproducible patterns of *lacZ* expression, whereas embryos bearing Region 7 produced no *lacZ* expression at the time points examined. Region 6 was not propagatible in bacteria and therefore not tested for regulatory activity. The white arrowhead points to the proepicardial organ (PEO) (C–D), while the black arrow indicates the dorsal aorta and the open arrow points to lacZ stained cells within the medial aspect of the developing heart (E).

portal (marked by a black chevron). It also produced strong expression in the PEO at E9.5, and at E11.5 additional domains of staining appeared in developing forelimb and hindlimb buds. These temporal and tissue specific differences in the patterns of reporter expression between Region 4 and 4A imply that elements positioned within the 3′ part of Region 4 play a role in modulating the enhancer readout of the 5′ components of this region.

The HGE reporter line carrying the enhancer spanning the 3′ end of Region 3 and 4A displayed a highly dynamic pattern of expression (Fig. 5). At E8.5 there was no expression. Discrete patches of cells flanking the developing heart appeared at E9.5 and these coalesced into groups of cells at E10.5. These *lacZ* expressing cells were present in the visceral endoderm, visceral mesoderm and within the developing E9.5 heart under the right side of the loop. The clonal populations of β-galactosidase marked cells were reminiscent of the knock-in of *lacZ* into the α-cardiac actin gene (Meilhac et al., 2003) which produced patches of *lacZ*-stained clones in the myocardium of the developing heart. There was a dramatic expansion of reporter staining at E11.5. Robust staining was seen in the ectoderm and endoderm, as well as in the surrounding epicardium of the heart and the dorsal aspect of the OFT. This resembles the pattern of expression produced by Region 3 at E9.5, except it had an anterior border of expression that incorporated the first branchial arch and surrounding tissues.

Together the results of the isolated enhancers and deleted variations suggest complex interactions are at play in this regulatory landscape flanking the *HoxB* complex. While some fragments in the enhancer scanning assay did not display regulatory activity, they may contain elements that cooperate with the components that we have uncovered to potentiate regulatory activity. This type of context dependent interaction both within and between these regulatory modules is reminiscent of a “holo-enhancer” described for *Fgf8* (Marinic et al., 2013).

The newly identified control regions function as shadow enhancers

The results from the 10 kb scanning analysis of regulatory potential indicate that a number of previously unidentified *cis*-acting modules reside within the area defined by Regions 2–4. In light of the difficulties we encountered in attempting to further

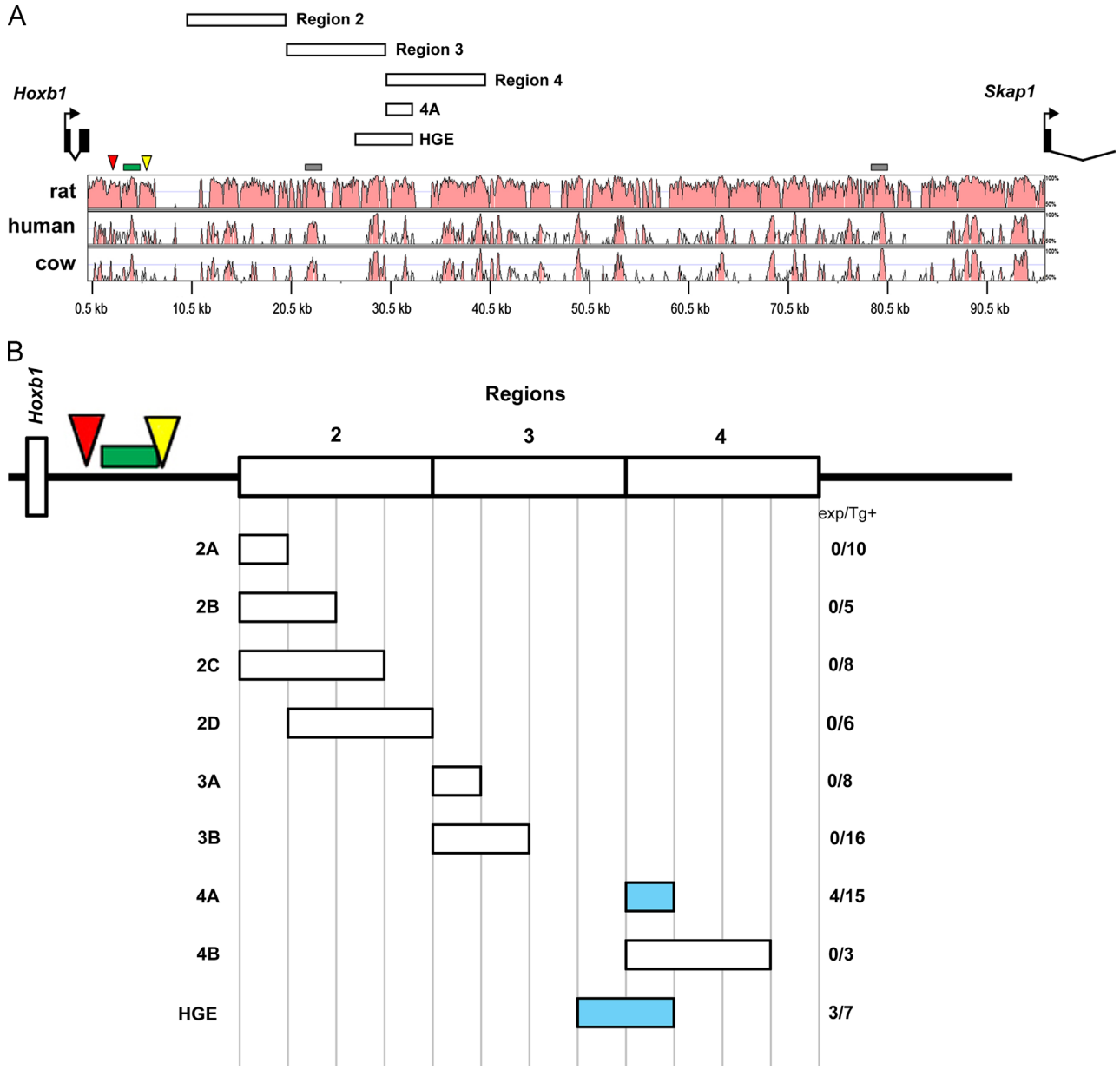


Fig. 4. Further reduction of the 10 kb subclones into smaller regions abolished reporter activity. (A) Comparison of the 10 kb reporters that had reproducible *lacZ* expression patterns (excluding Region 1) against a subset of the original *Hoxb1-Skap1* VISTA analysis shown in Fig. 1 reveals that these regions contained blocks of sequence conservation. (B) A series of successive 3′ truncations of Regions 2–4 were generated, but the majority of them failed to produce *lacZ* expression. Only subclone 4A produced *lacZ* expression (Fig. 5) but had different dynamics that the pattern produced by Region 4 from which it was derived. VISTA analysis indicated that the 3′ end of Region 3 contained a large block of sequence conservation. This region was added to 4A to produce the HGE subclone which produced *lacZ* expression from E9.5 onwards (Fig. 5).

isolate and characterize discrete or individual modular regulatory components contained within these regions we felt it was important to assess what contributions they made to the overall pattern of expression observed in the transposon-*lacZ* BAC reporter assay. Therefore, we generated an additional set of BAC transposon reporters, that deleted these newly identified regions in the presence or absence of the three previously identified enhancers 3′ of *Hoxb1*, using TI-1 as the base vector (Fig. 6). Removal of Regions 2–4 (TI-1 Δ2–4), produced a pattern of *lacZ* expression that closely resembled the original TI-1 BAC construct (Fig. 6). This would be consistent with the idea that this pattern can be driven by the known 3′ mesoderm and endoderm enhancers of the *Hoxb1* gene. However, when both the known 3′ *Hoxb1* enhancers and Regions 2–4 are deleted from this BAC transgenic reporter (TI-1Δ3′ Δ2–4), expression was completely abolished. These results underscore the contributions of Regions 2–4 and reveal a considerable degree of overlap with respect to the regulatory potential of both the local 3′ elements and those contained within Regions 2–4. Each appears capable of compensating for the loss of the other, but when both are deleted from the regulatory landscape adjacent to the *HoxB* complex, reporter expression is abolished. Therefore, Regions 2–4 may function as shadow enhancers with overlapping activities to the local *cis*-acting elements of *Hoxb1* which together could confer robustness in regulatory input (Hong et al., 2008; Perry et al., 2010). Alternatively, all of these regions could be working in concert as part of a larger "holo-enhancer", as has been proposed for the *Fgf8* gene (Marinic et al., 2013).

RAREs in the enhancers

In view of the important role for retinoids in heart development and modulation of *Hox* gene expression, we performed

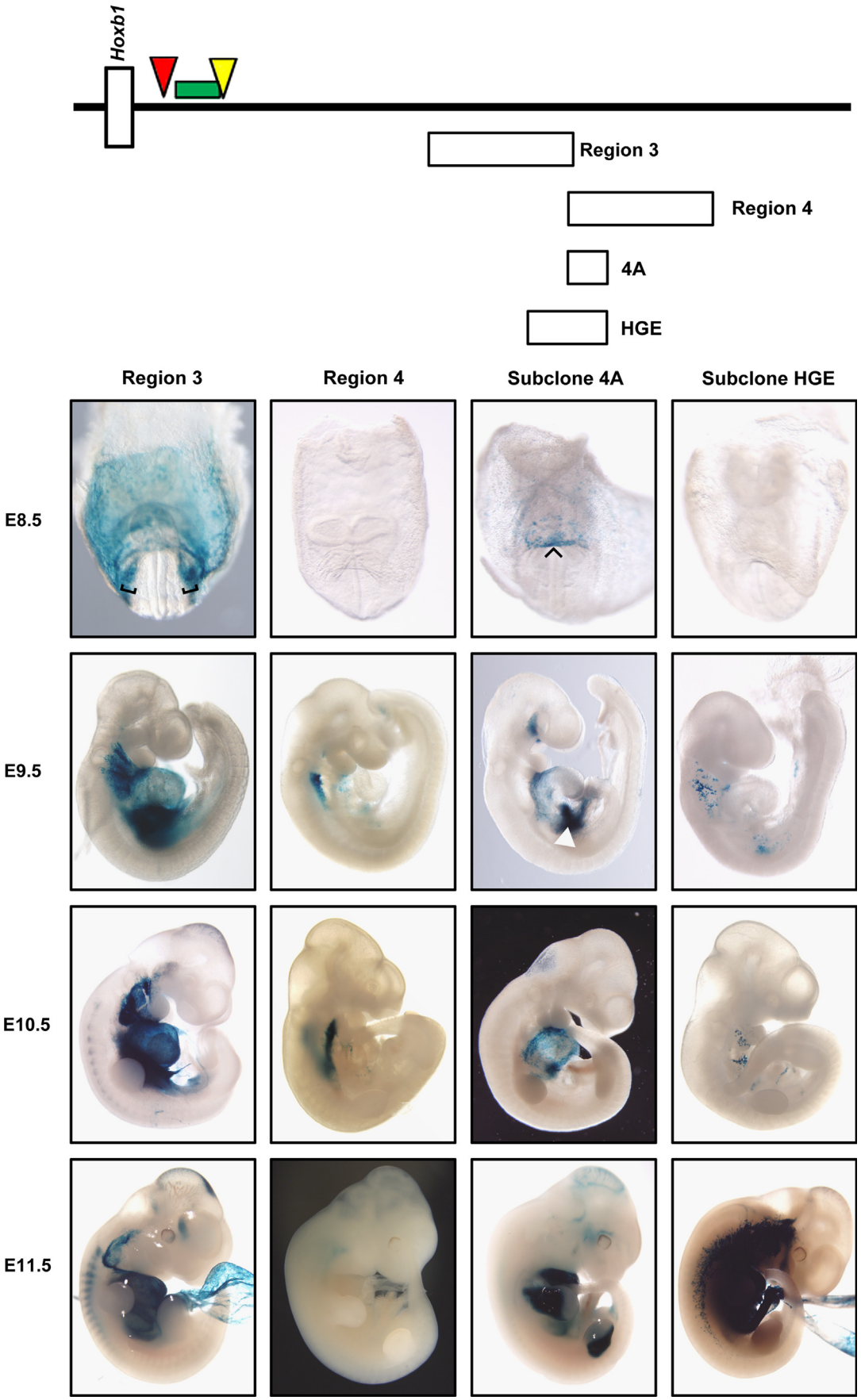


Fig. 5. Several regions downstream of the *HoxB* complex had dynamic expression patterns in the endoderm and tissues contributing to the developing heart. β-Galactosidase activity for Regions 3 and 4, and their derivatives (subclone 4A and subclone HGE) were examined from E8.5 to E11.5. At E8.5, Region 3 produced lacZ expression in the SHF (indicated by black, square brackets) while subclone 4A produced expression in the anterior intestinal portal (indicated by a black chevron).

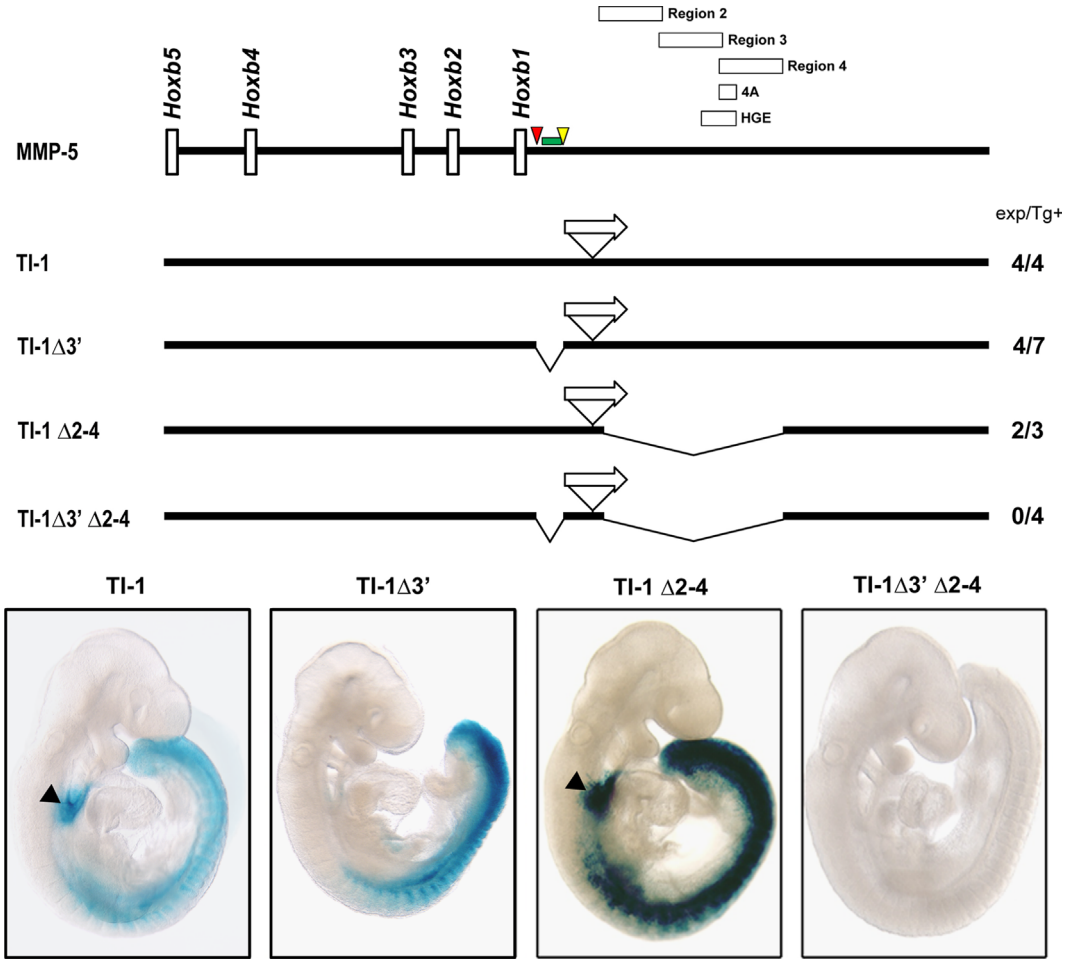


Fig. 6. The regions with novel enhancer activity function as shadow enhancers to the local *Hoxb1* enhancers. The original TI-1 reporter underwent three modifications. In the first modification, the local 3′ enhancers of *Hoxb1* were deleted (as previously shown in Fig. 2). In the second modification an area encompassing Regions 2–4 was deleted (TI-1 Δ2–4). In the final modification, both the local 3′ enhancers and the area encompassing Region 2–4 were deleted (TI-1Δ3′ Δ2–4). Deletion of both these regions abolished reporter expression. The black arrowhead indicated the foregut domain of expression driven by the local *Hoxb1* 3′ enhancers.

sequence analysis to search the entire intergenic region for the presence of RAREs. We detected two DR2 RAREs, one located in Region 3 and one in Region 4 (Fig. 7). The location of these RAREs raised the possibility that they contribute to the regulatory potential of these enhancers. Hence, we tested whether exogenously added RA alters reporter behavior in transgenic embryos carrying Region 3 and the HGE. Both Regions 3 and the HGE mediate a response to excess RA but the timing and type of response is different between the two (data not shown). We then mutated each putative RARE within these reporter constructs by converting the nucleotide sequences of the direct repeats to strings of adenines. Mutation of the DR2 RARE in Region 3 resulted in a significant decrease in reporter staining, while the mutation of the DR2 RARE in the HGE resulted in a greatly expanded pattern of reporter staining (Fig. 7). These findings suggest that RA signaling mediated through these RAREs plays a role in activity of the enhancers.

Analysis of the regulatory potential of the region flanking the HoxA complex

The *HoxA* and *HoxB* complexes arose by duplication and divergence from a common ancestor and the combined action of these duplicate clusters is specifically required for correct heart looping (Soshnikova et al., 2013). Therefore, based on the new heart and endodermal regulatory regions flanking *HoxB* identified above, we explored whether similar regulatory regions exist adjacent to the *HoxA* complex. To search for conservation between the intergenic regions flanking *HoxA* and *HoxB*, we performed a series of VISTA alignments (Fig. 8A). Comparison of the 3′ flanking intergenic regions from the mouse *HoxB* and *HoxA* complex revealed very few regions of sequence conservation (Fig. 8A). Looking specifically for similarities to Regions 2–4 of *HoxB*, we observed little if any sequence conservation. Sequence comparisons in placental mammals of the region flanking the *HoxA* complex itself shows a high degree of overall conservation spread throughout the 3′ intergenic region (Fig. 8B). This level of sequence conservation is not found when comparing the orthologous region of non-placental vertebrates (Fig. 8C). This is analogous to the type of overall conservation we observed in the intergenic region adjacent to *HoxB* in placental mammals (Fig. 1B). However, there is no significant overlap or similarity in the sequence profiles between the *HoxA* and *HoxB* clusters that points to the presence of potentially constrained regulatory regions shared by the sister clusters.

Since the regulatory activities of the regions flanking the *HoxB* complex in endoderm and heart tissues were located within approximately 40 kb of *Hoxb1*, we assayed the regulatory potential of similar regions flanking the *HoxA* complex by testing a series of 10 kb reporter constructs in transgenic mice (Fig. 9). Transgenic lines were generated with *HoxA* Regions 1, 3 and 4 but Region 2 was not propagable in bacteria (Fig. 10). Region 1 contains

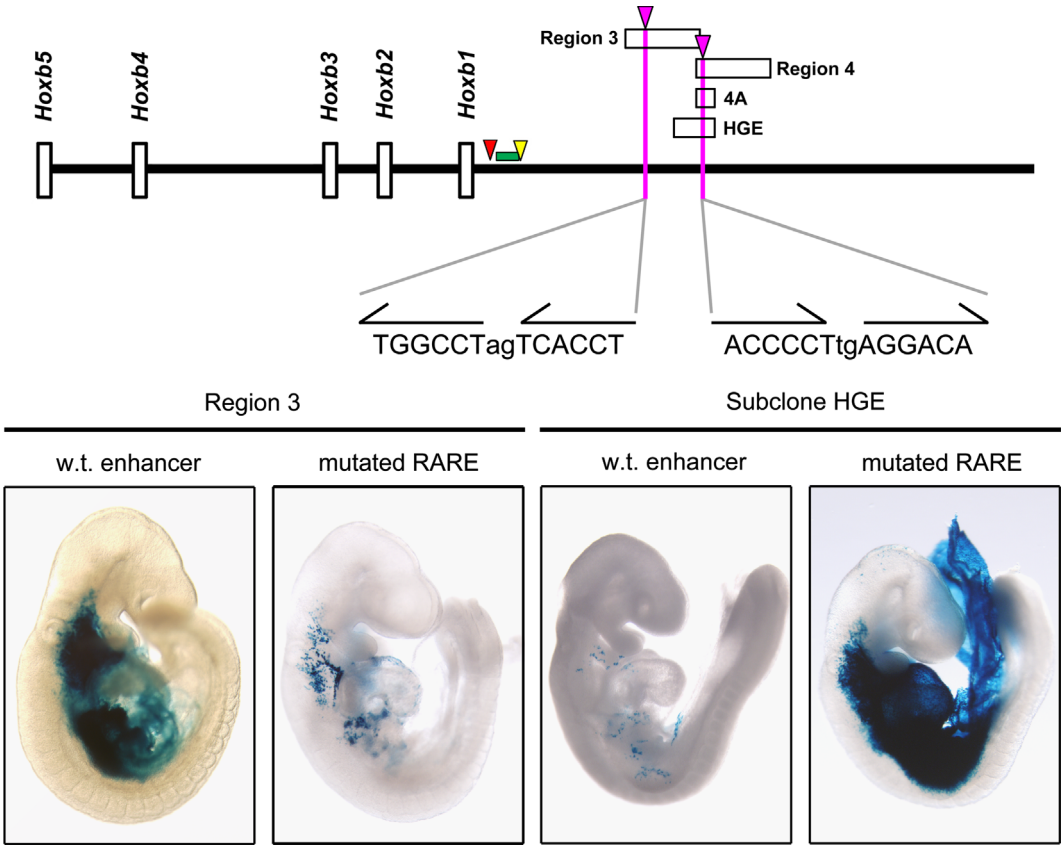


Fig. 7. Identification and testing of novel retinoic acid response elements (RAREs). Two novel DR2 RAREs within Regions 3 and 4 (subclone HGE) were identified based on sequence. Their approximate positions within the BAC are displayed as vertical purple bars. Below the BAC map, the sequence of each DR2 RARE is indicated. Mutation of these putative RAREs within these reporters either diminishes (Region 3 RARE) or increases (HGE subclone RARE) *lacZ* expression.

a previously characterized RARE-dependent neural enhancer required for early *Hoxa1* expression (Dupé et al., 1997). This reporter produced a pattern of *lacZ* expression in ectoderm and mesoderm at E8.5 very similar to endogenous *Hoxa1* expression, indicating that a mesodermal enhancer also resides within Region 1 (Fig. 10). However at later stages (E9.5–10.5) reporter expression stays on in the trunk and appears in forebrain regions at a time when endogenous *Hoxa1* is no longer expressed. This suggests that additional components or context within the flanking region may serve to restrict the activity of Region 1. Regions 3 and 4 adjacent to *HoxA* produced no *lacZ* expression from E8.5 to E9.5, however widespread expression was observed with Region 4 at E10.5 (Figs. 9 and 10). Hence, in addition to the lack of sequence similarity between intergenic regions adjacent to *HoxA* and *HoxB,* in *HoxA* there does not appear to be functionally equivalent *cis*-regulatory elements in a position similar to Regions 3 and 4 flanking *HoxB*.

We also scanned the *HoxA-Skap2* intergenic region for search for novel RARE sequences and identified four potential elements (Fig. 9B, purple arrowheads above the VISTA profile). Three of these RAREs are clustered in a region spanning the 3′ end of Region 1 and the 5′ end of Region 2. This cluster of RAREs downstream of the *Hoxa1* gene suggested that some type of regulatory potential may reside in this region. Since the entire *HoxA* Region 2 was not propagable in bacteria, we generated a smaller subclone (*HoxA* Region 2A) that contained the three clustered RAREs (Fig. 9). This construct produced strong *lacZ* expression throughout the endoderm and mesoderm from E8.5 to E10.5 (Fig. 10). Furthermore, when the heart was dissected at E10.5, *lacZ* expression domains clearly marked SHF derivatives, including the OFT and atria (Fig. 9). These regions marked by the *HoxA* Region 2A reporter in the developing heart corresponded to territories previously identified as being populated by *Hoxa1* and *Hoxa3*-expressing progenitors from the SHF (Bertrand et al., 2011). A transverse section of an E10.5 embryo expressing the *HoxA* Region 2A reporter showed that *lacZ* expression was mainly restricted to the myocardial layer of the heart (Fig. 9).

In summary, these analyses have shown that the 3′ flanking regions adjacent to *HoxA* and *HoxB* have the regulatory potential to modulate expression in endoderm and heart tissues, which may be associated with the functional role of these two clusters in heart looping. While the enhancers all have RAREs and may integrate aspects of retinoid signaling, the regions flanking the *HoxA* and *HoxB* complex do not appear to have a similar complement of conserved and ordered *cis*-regulatory elements spread throughout the intergenic regions which might be representative of heart or endodermal control regions present prior to duplication of these sister clusters.

Discussion

Analyses of the expression and function of *Hox* genes has demonstrated that they play an important role in modulating aspects of the cardiac developmental program (Bertrand et al., 2011; Chisaka and Capecchi, 1991; Makki and Capecchi, 2012; Soshnikova et al., 2013). However, very little was known about the *cis*-regulation that underlies these patterns of expression and function. In this study, we present regulatory analyses of the *HoxA* and *HoxB* complexes revealing that both contain multiple enhancers in their 3′ flanking intergenic regions capable of directing expression in endoderm and heart tissues over a long

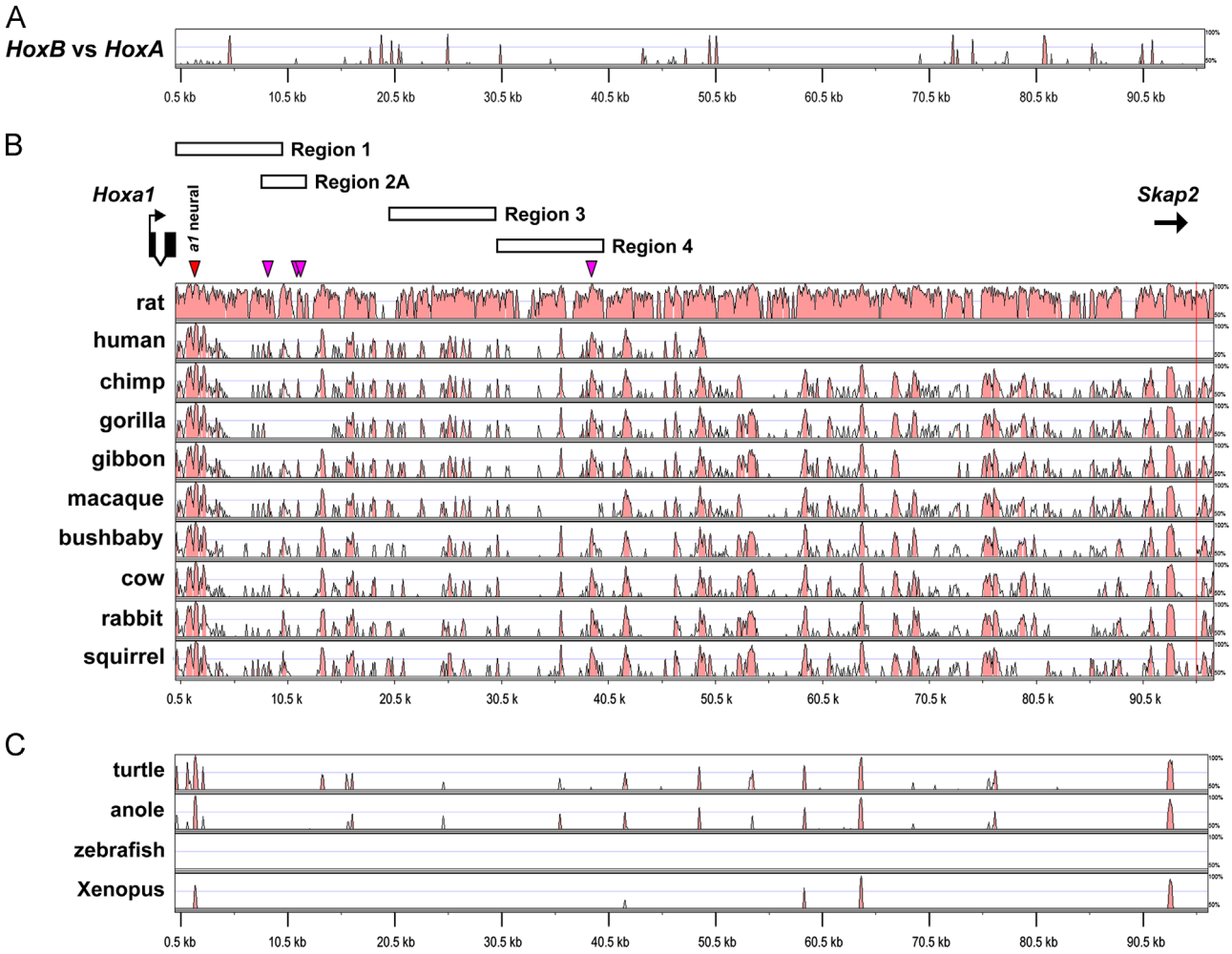


Fig. 8. The VISTA profile of the intergenic region flanking the *HoxA* complex is different than the profile of the paralogous region flanking the *HoxB* complex. (A) Comparison of the mouse intergenic regions flanking the *HoxB* and *HoxA* complexes reveal few areas of sequence conservation. (B) VISTA analysis of the intergenic region flanking the *HoxA* complex of placental mammals reveals multiple blocks of sequence conservation. Displayed above this VISTA profile are constructs generated for *lacZ* reporter analysis in transgenic assays (Fig. 9). Displayed about the VISTA profile is the known *Hoxa1* neural RARE (red triangle) as well as four purple triangles corresponding to potentially novel RAREs. Three of these elements reside within the 2A region. (C) The intergenic region flanking the *HoxA* complex has poor sequence conservation when compared against non-placental vertebrates.

range. The control modules in the *HoxB* complex are spread throughout the flanking intergenic region, display complex organization and have some degree of overlapping activity with local enhancers suggesting that they may function as shadow enhancers. These enhancers have RAREs implying that retinoid signaling serves to directly integrate *Hox* gene expression with cardiogenesis. Together our findings provide insight for understanding features of the regulatory landscape of *Hox* clusters and flanking intergenic regions during vertebrate evolution with respect to their roles in endoderm and heart development. These data support a regulatory model whereby the conserved control modules flanking the *HoxA* and *HoxB* clusters act over a long range on multiple *Hox* genes to generate their nested patterns of expression during cardiogenesis.

Novel endoderm and heart enhancers

Our analyses have uncovered multiple regions that displayed discrete localized *cis*-regulatory activity in endoderm and heart tissues. When the 3′ flanking region of the *HoxB* complex was broken up into 10 kb fragments, enhancer activity for the visceral endoderm, visceral mesoderm and septum transversum were segregated away from that of the presomitic and somitic mesoderm. Only Region 1 produced *lacZ* expression in these latter domains, whereas Regions 2–4 produced novel patterns of *lacZ* expression in the PEO, epicardium, visceral endoderm and visceral mesoderm (Fig. 3). The enhancers located at the 3′ end of the *HoxA* complex also mediated reporter staining in PEO, visceral endoderm and visceral mesoderm at E.9.5 and in SHF derivatives (OFT and atria) at E10.5. The *HoxA* Regions 1 and 2A enhancers also direct reporter expression in a variety of other tissues, including neural ectoderm and mesoderm (Figs. 9 and 10). These patterns are similar to those mediated by the three previously characterized *Hoxb1* enhancers (Region 1) located in a similar position at the 3′ end of the *HoxB* complex. Based on position and activity there seems to be some commonality between the regulatory regions immediately flanking the *HoxA* and *HoxB* complexes. However, *HoxB* has additional enhancers spread throughout the more distal intergenic region.

Regulatory complexity and long range activity of flanking enhancers

We attempted to break these enhancers down into smaller modules by deletion analysis to facilitate the characterization of *cis*-regulatory elements and transcription factor binding motifs which potentiate their activities. Unlike the modular enhancers previously characterized for *Hox* gene expression in neural and mesodermal tissues (Alexander et al., 2009; Tumpel et al., 2009),

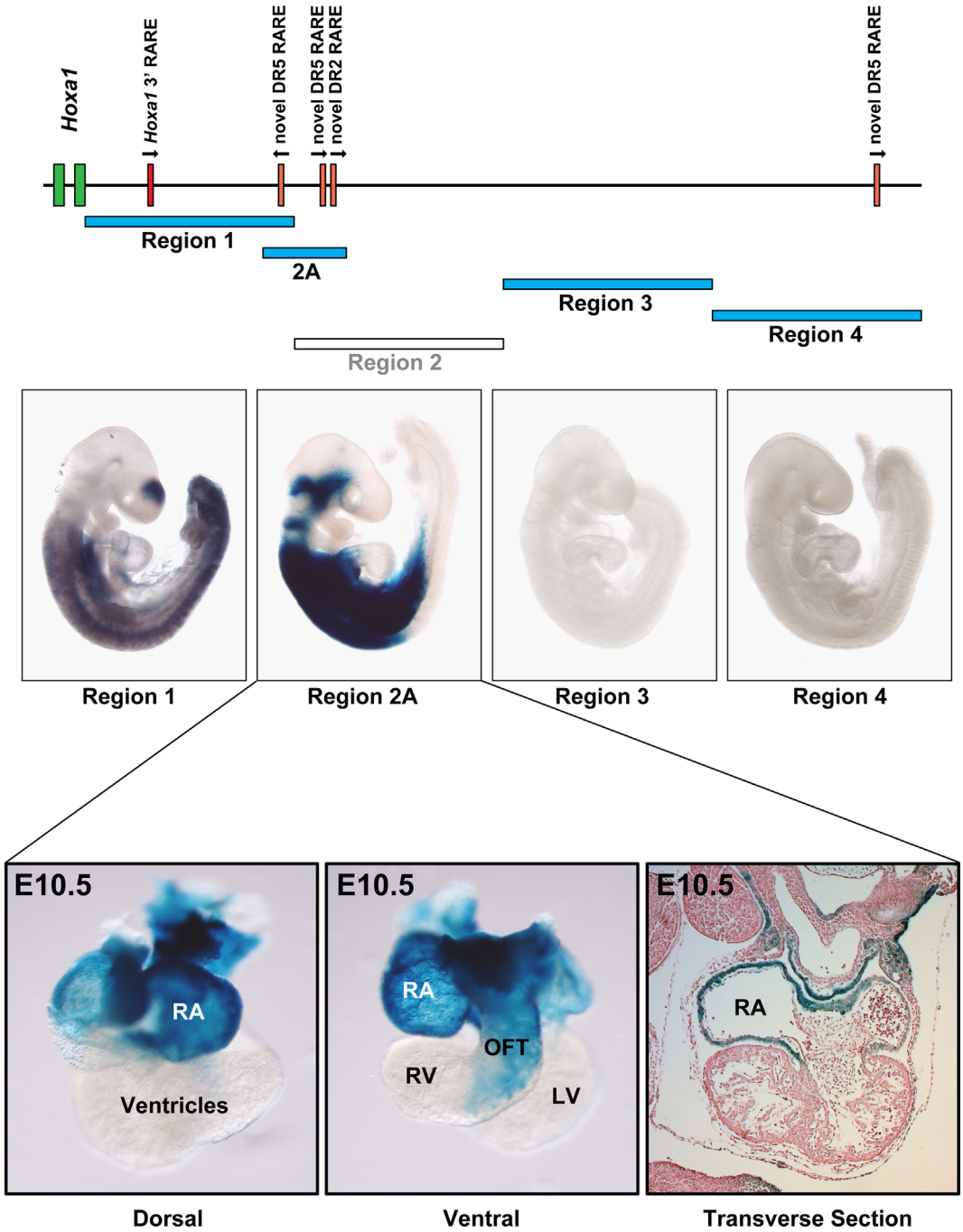


Fig. 9. Transgenic analyses at E9.5 of the regions flanking the *HoxA* complex show that they have different regulatory potentials. *HoxA's* Region 1 contains the *Hoxa1* neural RARE and produces expression throughout the trunk as well as in the forebrain. *HoxA* Regions 3 and 4 produced no reporter activity at E9.5. *HoxA* Region 2 was not tested, but a subclone of it (2A) containing a portion of the 3′ end of Region 1 produced robust expression throughout most of the trunk mesoderm and endoderm as well as in the developing heart. In the E10.5 heart, the 2A region produced *lacZ* expression in many of the SHF derivatives of the heart. A transverse section through the E10.5 transgenic heart shows that *lacZ* expression of *HoxA* Region 2A reporter was restricted primarily to the myocardium.

further reduction of the enhancers identified in this study generally led to either a loss or change in activity (Fig. 4). For example, *HoxB* Region 4 generated *lacZ* expression in the dorsal aorta, whereas the 4A 5′ subfragment produced expanded reporter staining in the epicardium (Fig. 5). Collectively these results imply that multiple *cis*-elements spread throughout the 10 kb enhancer regions work in combination to modulate regulatory activity.

Consistent with this type of complexity, some of the isolated individual enhancers display regulatory activity different from that observed when they were in the full context of the *HoxB* BAC carrying the transposon reporters. This is illustrated by *HoxB* Region 3, which as an isolated enhancer directs robust *lacZ* expression in the visceral ectoderm as well as in the lateral mesoderm (Figs. 3D and 5). However, in the context of the BAC, TI-2 and TI-3 display more restricted domains of expression in these tissues (Fig. 2B). Therefore, the enhancers in Region 2–4 flanking *HoxB* cluster appear to be dependent upon context to properly mediate the appropriate domains of expression.

In the *HoxA* analysis, Regions 1 and 2A directed reporter expression at E8.5 with anterior boundaries and tissue specificity that closely resemble the pattern of the endogenous *Hoxa1* gene. However, at later stages reporter expression persists and appears in

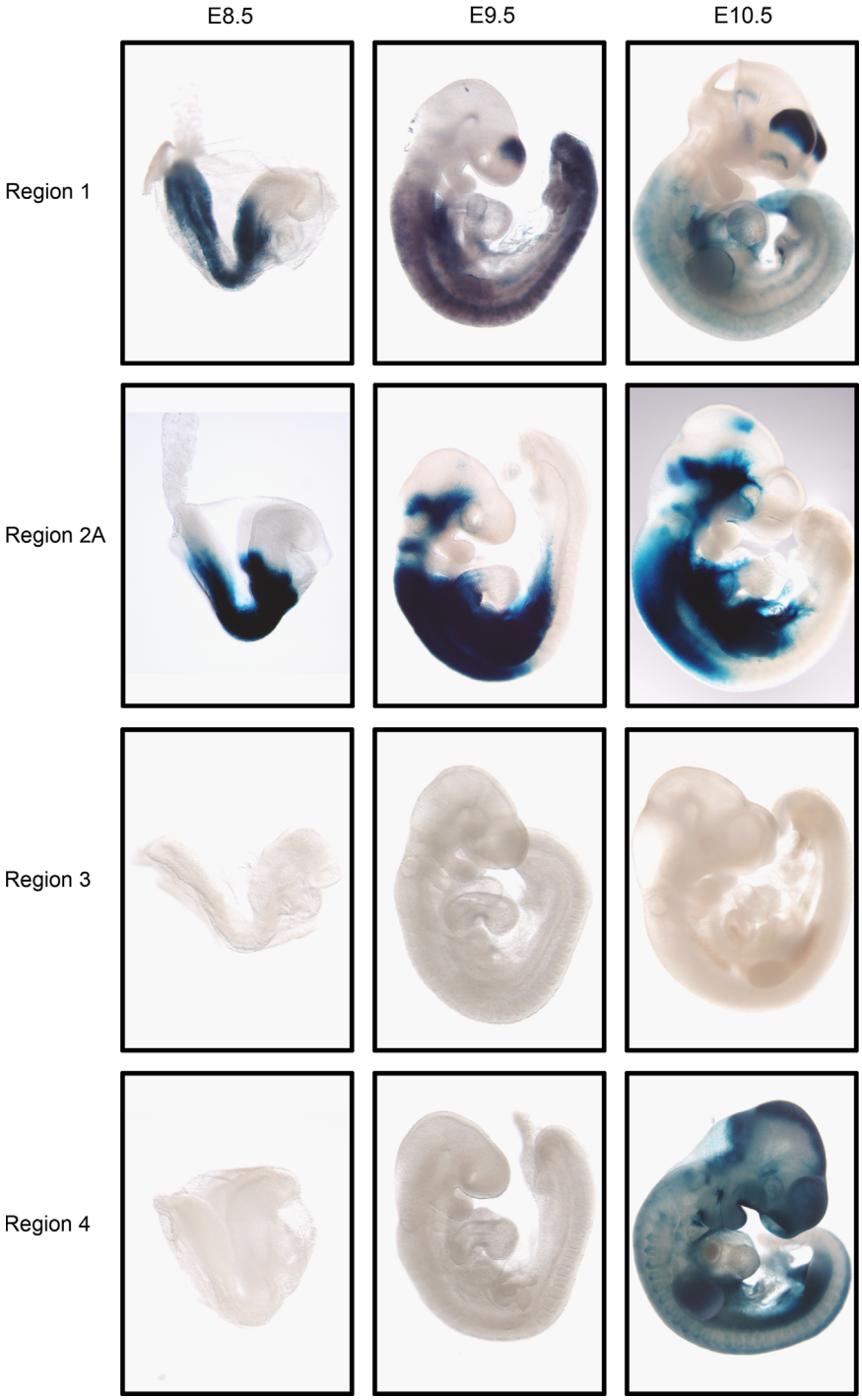


Fig. 10. Temporal time course of the enhancer potential of the *HoxA* regions in transgenic mouse embryos from E8.5 to E10.5. *HoxA's* Region 1 produced expression throughout the trunk mesoderm and endoderm from E8.5 to E10.5, with *lacZ* intensity decreasing at E10.5. Region 2A produced robust *lacZ* expression throughout the trunk mesoderm and endoderm from E8.5 to E10.5, but became progressively excluded from the developing posterior trunk. Region 3 produced no *lacZ* expression at all stages examined. Region 4 had no early enhancer activity, producing *lacZ* expression only in E10.5 embryos throughout defined regions that included neural tissues, rostral somites and visceral endoderm.

ectopic domains (Fig. 10) indicating that both the pattern and timing of these regulatory regions no longer correlate with *Hoxa1* expression. This suggests that either additional elements or the context within the *HoxA* flanking region maybe important in restricting the pattern and timing of these regulatory regions.

The inability to break these regions down into modular units and their context dependence suggest that these enhancers are part of an extended regulatory landscape necessary to properly potentiate their activity. Conceptually this could be considered to be a holo-enhancer analogous to that described for the *Fgf8* gene (Marinic et al., 2013). In this model, the output of multiple enhancers spread over a broad region depends upon their position or context to regulate the intrinsic activities of the regulatory components as a coherent unit. This type of context dependent activity could involve differential looping to produce dynamic and specific transcriptional readouts. By analyzing individual enhancers or reducing them into smaller fragments in our assays, the local chromatin architecture maybe disrupted resulting in different regulatory outcomes.

It is an attractive possibility to think that collectively these 3′ flanking enhancers could function globally over the *HoxB* complex to direct expression of multiple *Hox* genes. This suggests a mechanism whereby long range interactions of these enhancers could be shared by multiple genes to create the endogenous domains of nested *Hox* expression in the SHF. We found evidence that some of these enhancers can work over long range. In the *HoxB* BAC reporter assays, the regulatory influence of the enhancers extended throughout the intergenic region towards the *Skap1* gene, as the *lacZ* reporter transposed at 10 kb, 21 kb, 26 kb, 52 kb and 64 kb downstream of the *Hoxb1* gene produced a similar pattern of expression (Fig. 2). In contrast, the *Hoxb1* 3′ endoderm enhancer has a limited range of influence, since it could only activate the *lacZ* reporter in TI-1 (4.4 kb distance) and fails to work on the more distal transposon reporters.

Shadow enhancers

Despite the context dependence and complexity of the organization of the enhancers flanking the *HoxB* complex, we uncovered regions that had a considerable degree of overlap in their ability to regulate expression. Furthermore, only deletion of both Regions 2–4 and the three previously identified enhancers 3′ of *Hoxb1* (Region 1) eliminated reporter activity (Fig. 6). The overlap in activity and functional compensation in BAC assays imply that Regions 2–4 may function as shadow enhancers to the local *cis*-acting enhancers 3′ of *Hoxb1*. In view of the dynamic regulation of *Hox* genes during these stages of cardiac development, the purpose of these shadow enhancers could be to ensure robustness and appropriate timing of *HoxB* expression in these tissues (Hong et al., 2008; Perry et al., 2010). Our experiments suggest that there may also be shadow enhancers for the presomitic mesoderm. Region 1 contains a mesodermal enhancer of *Hoxb1* (Marshall et al., 1994). However, deletion of this region in BAC assays still displayed reporter expression in the presomitic mesoderm indicating that other enhancers in Regions 2–4 contribute to the mesodermal expression profile (Fig. 6).

Upstream regulators: RA signaling

Despite the fact that these regulatory regions were broadly dispersed throughout the 3′ flanking intergenic region, we used the conservation and regulatory activity as a basis to search for potential upstream regulators that may potentiate enhancer activity. For both the *HoxA* and *B* clusters, RAREs appear to play an integral role in the regulatory activity of the enhancers. The two novel RAREs characterized in the *HoxB* flanking region appear to have different activities. The RARE in Region 3 correlates with a role in activation, as mutation of its RARE results in significant down-regulation in reporter activity (Fig. 7). Conversely, the RARE of Region 4/HGE is consistent with a role in repression as mutation of the RARE results in dramatic up-regulation in reporter staining. These differences are consistent with the multiple roles of RA signaling in establishing the precise boundaries of the cardiac fields (Buckingham et al., 2005; Vincent and Buckingham, 2010).

Evolution of regulatory features in HoxA and HoxB sister clusters

The maintenance of the clustered organization of *Hox* genes is believed to be in part due to regulatory constraints associated with features such as, colinearity, the action of long range global regulatory elements and enhancer sharing (Kmita and Duboule, 2003; Noordermeer and Duboule, 2013). The complex architecture of the regulatory landscape revealed by our analysis of flanking regions of the *HoxB* complex is consistent with this idea. If the shadow enhancers and context dependent organization of this broad intergenic region are essential to mediate the ordered expression of multiple *HoxB* members in endoderm and cardiac development, they would place a regulatory constraint on disrupting the complex and 3′ flanking region.

The specific functional requirement for the combined action of the *HoxA* and the *HoxB* complexes for correct heart looping, suggested that a novel *cis*-element or module present in the A/B ancestor cluster may have been maintained in the sister copies during duplication of the cluster in vertebrate evolution (Soshnikova et al., 2013). Using phylogenetic footprinting analyses in both placental and nonplacental vertebrates, we did not detect areas of sequence conservation which would be candidates for such an ancestral element. However, our regulatory analyses did highlight functional similarities between the regions immediately flanking the 3′ end of the *HoxA* and the *HoxB* clusters. The Region 1 fragments from both clusters produced very similar patterns of reporter expression in neural ectoderm, mesoderm and endoderm (Figs. 3B and 9). These regions contain neural enhancers with RAREs required for the expression of *Hoxa1* and *Hoxb1* (Dupé et al., 1997; Marshall et al., 1994; Studer et al., 1998). While it was known that a mesodermal and an endodermal enhancer were adjacent to the neural enhancer in *HoxB*, our analysis uncovered similar regulatory potential to modulate expression in mesoderm and endoderm in the corresponding region adjacent to the *HoxA*. Furthermore, we found a new cluster of RAREs within this *HoxA* regulatory region which may be analogous to the RARE found in the *HoxB* endoderm enhancer (Huang et al., 1998). While we did not find extensive sequence conservation between the *HoxA* and *HoxB* complexes in this region, the relative positions adjacent to the complexes, the functional/regulatory similarities and the presence of RAREs suggest that these may have evolved by duplication and divergence from a common ancestor. The absence of extended sequence conservation may be a consequence of the fact that the precise order and number of binding sites for key upstream factors in many enhancers can vary considerably during evolution (Moses et al., 2006).

The endodermal and heart enhancers we identified in Regions 2–4 in the *HoxB* complex do not appear to have equivalent counterparts flanking the *HoxA* complex, although we have not characterized the complete intergenic region between *HoxA* and *Skap2* which is much larger than that between *HoxB* and *Skap1*. Hence, the *HoxB* shadow enhancers appear to be specific to the *HoxB* complex, arising after duplication of the A/B sister complexes. Our results on the similarities in location, enhancer output, and dependence on retinoid signaling suggest a model in which the proximal 3′ enhancers adjacent to the *HoxA* and *HoxB* complex may have arisen from a common ancestor and play primary roles in endodermal and cardiac regulation of *Hox* genes. The shadow enhancers in Regions 2–4 of *HoxB* may then make a contribution by reinforcing the timing, robustness or adding new domains to these patterns of expression. Together our analyses suggest a common regulatory mechanism, whereby the conserved control modules act over a long range on multiple *Hox* genes to generate nested patterns of *HoxA* and *HoxB* expression during cardiogenesis essential for proper looping of the heart.

Acknowledgments

We thank members of the Krumlauf lab for valuable discussions, Mark Parrish and Youngwook Ahn for advice on BAC recombineering, the Stowers Institute Histology facility and Heidi Monnin, Andrea Moran and other members of the Stowers Laboratory Animal Services Facility for their care of our mice and technical assistance. All experiments involving mice were approved by the Institutional Animal Care and Use Committee of the Stowers Institute for Medical Research (Protocol 2010-0062). These experiments were initiated at the National Institute for

Medical Research (U.K.) under support by the Medical Research Council provided to T.J., M.T.M.P. and R.K.. T.J. was also supported in part by a Ruth Kirkstein NRSA Postdoctoral Fellowship (#HD008457). This research and C.D.N., X.W. and R.K. were supported by funds from the Stowers Institute.

References

Ahn, Y., Sims, C., Logue, J.M., Weatherbee, S.D., Krumlauf, R., 2013. Lrp4 and Wise interplay controls the formation and patterning of mammary and other skin appendage placodes by modulating Wnt signaling. Development 140, 583–593.

Alexander, T., Nolte, C., Krumlauf, R., 2009. Hox genes and segmentation of the hindbrain and axial skeleton. Annu. Rev. Cell Dev. Biol. 25, 431–456.

Aubin, J., Dery, U., Lemieux, M., Chailler, P., Jeannotte, L., 2002. Stomach regional specification requires Hoxa5-driven mesenchymal–epithelial signaling. Development 129, 4075–4087.

Bertrand, N., Roux, M., Ryckebusch, L., Niederreither, K., Dolle, P., Moon, A., Capecchi, M., Zaffran, S., 2011. Hox genes define distinct progenitor subdomains within the second heart field. Dev. Biol. 353, 266–274.

Boulet, A.M., Capecchi, M.R., 1996. Targeted disruption of*Hoxc4* causes esophageal defects and vertebral transformations. Dev. Biol. 177, 232–249.

Buckingham, M., Meilhac, S., Zaffran, S., 2005. Building the mammalian heart from two sources of myocardial cells. Nat. Rev. Genet. 6, 826–835.

Chisaka, O., Capecchi, M.R., 1991. Regionally restricted developmental defects resulting from targeted disruption of the mouse homeobox gene *hox1.5*. Nature 350, 473–479.

Creazzo, T.L., Godt, R.E., Leatherbury, L., Conway, S.J., Kirby, M.L., 1998. Role of cardiac neural crest cells in cardiovascular development. Annu. Rev. Physiol. 60, 267–286.

Deschamps, J., van Nes, J., 2005. Developmental regulation of the Hox genes during axial morphogenesis in the mouse. Development 132, 2931–2942.

Diman, N.Y., Remacle, S., Bertrand, N., Picard, J.J., Zaffran, S., Rezsohazy, R., 2011. A retinoic acid responsive Hoxa3 transgene expressed in embryonic pharyngeal endoderm, cardiac neural crest and a subdomain of the second heart field. PLoS One 6, e27624.

Dollé, P., Izpisùa-Belmonte, J.C., Brown, J.M., Tickle, C., Duboule, D., 1991. *Hox-4* genes and the morphogenesis of mammalian genitalia. Genes Dev. 5, 1767–1776.

Duboule, D., Dollé, P., 1989. The structural and functional organization of the murine *HOX* gene family resembles that of *Drosophila* homeotic genes. EMBO J. 8, 1497–1505.

Dunwoodie, S.L., 2007. Combinatorial signaling in the heart orchestrates cardiac induction, lineage specification and chamber formation. Semin. Cell Dev. Biol. 18, 54–66.

Dupé, V., Davenne, M., Brocard, J., Dollé, P., Mark, M., Dierich, A., Chambon, P., Rijli, F., 1997. *In vivo* functional analysis of the *Hoxa1* 3′ retinoid response element (3′ RARE). Development 124, 399–410.

Gavalas, A., Studer, M., Lumsden, A., Rijli, F.M., Krumlauf, R., Chambon, P., 1998. *Hoxa1* and *Hoxb1* synergize in patterning the hindbrain, cranial nerves and second pharyngeal arch. Development 125, 1123–1136.

Graham, A., Papalopulu, N., Krumlauf, R., 1989. The murine and *Drosophila* homeobox gene complexes have common features of organization and expression. Cell 57, 367–378.

Grapin-Botton, A., 2005. Antero-posterior patterning of the vertebrate digestive tract: 40 years after Nicole Le Douarin's PhD thesis. Int. J. Dev. Biol. 49, 335–347.

Ho, S.N., Hunt, H.D., Horton, R.M., Pullen, J.K., Pease, L.R., 1989. Site-directed mutagenesis by overlap extension using the polymerase chain reaction. Gene 77, 51–59.

Hong, J.W., Hendrix, D.A., Levine, M.S., 2008. Shadow enhancers as a source of evolutionary novelty. Science 321, 1314.

Huang, D., Chen, S.W., Langston, A.W., Gudas, L.J., 1998. A conserved retinoic acid responsive element in the murine *Hoxb-1* gene is required for expression in the developing gut. Development 125, 3235–3246.

Keegan, B.R., Feldman, J.L., Begemann, G., Ingham, P.W., Yelon, D., 2005. Retinoic acid signaling restricts the cardiac progenitor pool. Science 307, 247–249.

Kinkel, M.D., Sefton, E.M., Kikuchi, Y., Mizoguchi, T., Ward, A.B., Prince, V.E., 2009. Cyp26 enzymes function in endoderm to regulate pancreatic field size. Proc. Natl. Acad. Sci. USA 106, 7864–7869.

Kmita, M., Duboule, D., 2003. Organizing axes in time and space; 25 years of colinear tinkering. Science 301, 331–333.

Kondo, T., Dollé, P., Zakany, J., Duboule, D., 1996. Function of posterior HoxD genes in the morphogenesis of the anal sphincter. Development 122, 2651–2659.

Krumlauf, R., 1994. *Hox* genes in vertebrate development. Cell 78, 191–201.

Lee, E.C., Yu, D., Martinez de Velasco, J., Tessarollo, L., Swing, D.A., Court, D.L., Jenkins, N.A., Copeland, N.G., 2001. A highly efficient *Escherichia coli*-based chromosome engineering system adapted for recombinogenic targeting and subcloning of BAC DNA. Genomics 73, 56–65.

Makki, N., Capecchi, M.R., 2012. Cardiovascular defects in a mouse model of HOXA1 syndrome. Hum. Mol. Genet. 21, 26–31.

Mallo, M., Wellik, D.M., Deschamps, J., 2010. Hox genes and regional patterning of the vertebrate body plan. Dev. Biol. 344, 7–15.

Manley, N., Capecchi, M., 1998. Hox group 3 paralogs regulate the development and migration of the thymus, thyroid and parathyroid glands. Dev. Biol. 195, 1–15.

Marinic, M., Aktas, T., Ruf, S., Spitz, F., 2013. An integrated holo-enhancer unit defines tissue and gene specificity of the Fgf8 regulatory landscape. Dev. Cell 24, 530–542.

Marshall, H., Studer, M., Pöpperl, H., Aparicio, S., Kuroiwa, A., Brenner, S., Krumlauf, R., 1994. A conserved retinoic acid response element required for early expression of the homeobox gene *Hoxb-1*. Nature 370, 567–571.

Meilhac, S.M., Kelly, R.G., Rocancourt, D., Eloy-Trinquet, S., Nicolas, J.F., Buckingham, M.E., 2003. A retrospective clonal analysis of the myocardium reveals two phases of clonal growth in the developing mouse heart. Development 130, 3877–3889.

Morgan, B.A., Conlon, F.L., Manzanares, M., Millar, J.B.A., Kanuga, N., Sharpe, J., Krumlauf, R., Smith, J.C., Sedgwick, S.G., 1996. Transposon tools for recombinant DNA manipulation: characterisation of transcriptional regulators from yeast, xenopus and mouse. Proc. Natl. Acad. Sci. 93, 2801–2806.

Moses, A.M., Pollard, D.A., Nix, D.A., Iyer, V.N., Li, X.Y., Biggin, M.D., Eisen, M.B., 2006. Large-scale turnover of functional transcription factor binding sites in Drosophila. PLoS Comput. Biol. 2, e130.

Niederreither, K., Subbarayan, V., Dollé, P., Chambon, P., 1999. Embryonic retinoic acid synthesis is essential for early mouse post-implantation development. Nat. Genet. 21, 444–448.

Niederreither, K., Vermot, J., Messaddeq, N., Schuhbaur, B., Chambon, P., Dolle, P., 2001. Embryonic retinoic acid synthesis is essential for heart morphogenesis in the mouse. Development 128, 1019–1031.

Noordermeer, D., Duboule, D., 2013. Chromatin architectures and Hox gene collinearity. Curr. Top. Dev. Biol. 104, 113–148.

Parrish, M., Unruh, J., Krumlauf, R., 2011. BAC modification through serial or simultaneous use of CRE/Lox technology. J. Biomed. Biotechnol. 2011, 924068.

Perry, M.W., Boettiger, A.N., Bothma, J.P., Levine, M., 2010. Shadow enhancers foster robustness of Drosophila gastrulation. Curr. Biol. 20, 1562–1567.

Rhinn, M., Dolle, P., 2012. Retinoic acid signalling during development. Development 139, 843–858.

Ryckebusch, L., Wang, Z., Bertrand, N., Lin, S.C., Chi, X., Schwartz, R., Zaffran, S., Niederreither, K., 2008. Retinoic acid deficiency alters second heart field formation. Proc. Natl. Acad. Sci. USA 105, 2913–2918.

Searcy, R.D., Yutzey, K.E., 1998. Analysis of *Hox* gene expression during early avian heart development. Dev. Dyn. 213, 82–91.

Sirbu, I.O., Gresh, L., Barra, J., Duester, G., 2005. Shifting boundaries of retinoic acid activity control hindbrain segmental gene expression. Development 132, 2611–2622.

Soshnikova, N., Dewaele, R., Janvier, P., Krumlauf, R., Duboule, D., 2013. Duplications of hox gene clusters and the emergence of vertebrates. Dev. Biol. 378, 194–199.

Stafford, D., White, R.J., Kinkel, M.D., Linville, A., Schilling, T.F., Prince, V.E., 2006. Retinoids signal directly to zebrafish endoderm to specify insulin-expressing beta-cells. Development 133, 949–956.

Stoller, J.Z., Epstein, J.A., 2005. Cardiac neural crest. Semin. Cell Dev. Biol. 16, 704–715.

Studer, M., Gavalas, A., Marshall, H., Ariza-McNaughton, L., Rijli, F., Chambon, P., Krumlauf, R., 1998. Genetic interaction between *Hoxa1* and *Hoxb1* reveal new roles in regulation of early hindbrain patterning. Development 125, 1025–1036.

Tischfield, M.A., Bosley, T.M., Salih, M.A., Alorainy, I.A., Sener, E.C., Nester, M.J., Oystreck, D.T., Chan, W.M., Andrews, C., Erickson, R.P., Engle, E.C., 2005. Homozygous HOXA1 mutations disrupt human brainstem, inner ear, cardiovascular and cognitive development. Nat. Genet. 37, 1035–1037.

Tumpel, S., Wiedemann, L.M., Krumlauf, R., 2009. Hox genes and segmentation of the vertebrate hindbrain. Curr. Top. Dev. Biol. 88, 103–137.

Tümpel, S., Wiedemann, L.M., Krumlauf, R., 2009. Hox genes and segmentation of the vertebrate hindbrain. In: Pourquié, O. (Ed.), HOX Genes. Academic Press, Burlington, pp. 103–137.

Varner, V.D., Taber, L.A., 2012. Not just inductive: a crucial mechanical role for the endoderm during heart tube assembly. Development 139, 1680–1690.

Vincent, S.D., Buckingham, M.E., 2010. How to make a heart: the origin and regulation of cardiac progenitor cells. Curr. Top. Dev. Biol. 90, 1–41.

Wang, H., Rudd, C.E., 2008. SKAP-55, SKAP-55-related and ADAP adaptors modulate integrin-mediated immune-cell adhesion. Trends Cell Biol. 18, 486–493.

Waxman, J.S., Keegan, B.R., Roberts, R.W., Poss, K.D., Yelon, D., 2008. Hoxb5b acts downstream of retinoic acid signaling in the forelimb field to restrict heart field potential in zebrafish. Dev. Cell 15, 923–934.

Wellik, D.M., 2009. Hox genes and vertebrate axial pattern. Curr. Top. Dev. Biol. 88, 257–278.

Whiting, J., Marshall, H., Cook, M., Krumlauf, R., Rigby, P.W., Stott, D., Allemann, R.K., 1991. Multiple spatially specific enhancers are required to reconstruct the pattern of *Hox-2.6* gene expression. Genes Dev. 5, 2048–2059.

Wolgemuth, D.J., Behringer, R.R., Mostoller, M.P., Brinster, R.L., Palmiter, R.D., 1989. Transgenic mice overexpressing the mouse homoeobox-containing gene Hox-1.4 exhibit abnormal gut development. Nature 337, 464–467.

Yee, S.-P., Rigby, P.W.J., 1993. The regulation of *myogenin* gene expression during the embryonic development of the mouse. Genes Dev. 7, 1277–1289.

Yutzey, K.E., Rhee, J.T., Bader, D., 1994. Expression of the atrial-specific myosin heavy chain AMHC1 and the establishment of anteroposterior polarity in the developing chicken heart. Development 120, 871–883.

Zacchetti, G., Duboule, D., Zakany, J., 2007. Hox gene function in vertebrate gut morphogenesis: the case of the caecum. Development 134, 3967–3973.

Zakany, J., Duboule, D., 1999. Hox genes and the making of sphincters. Nature 401, 761–762.

Zakany, J., Duboule, D., 2007. The role of Hox genes during vertebrate limb development. Curr. Opin. Genet. Dev. 17, 359–366.

Zákány, J., Kmita, M., Alarcon, P., de la Pompa, J.L., Duboule, D., 2001. Localized and transient transcription of Hox genes suggests a link between patterning and the segmentation clock. Cell 106, 207–217.